Recurrent Graph Convolutional Network can learn the graph topologies across consecutive frames.

Recurrent Graph Convolutional Networks for Skeleton-based Action Recognitions

INTRO

- The skeleton can be represented by a graph, which each node represents one joint, and human actions can be viewed as spatio-temporal graphs.
- There is no natural connection between the two hands, but the interaction between them plays an important role.

METHODS

1. **Dot Product**: $E(X_t) = X_t^T X_t$,
2. **Gaussian**: $E(X_t) = e^{X_t^T X_t}$,
3. **Embedded Gaussian**: $E(X_t) = e^{\theta(X_t)^T \emptyset(X_t)}$,
4. **Latent connection**: $L_k^t = LSTM(E(W_k f_{in}^t), L_k^{t-1})$,
5. **Spatial graph convolution**: $f_{out} = \sum_k^{K_v} W_k f_{in}(A_k + L_k^t)$.

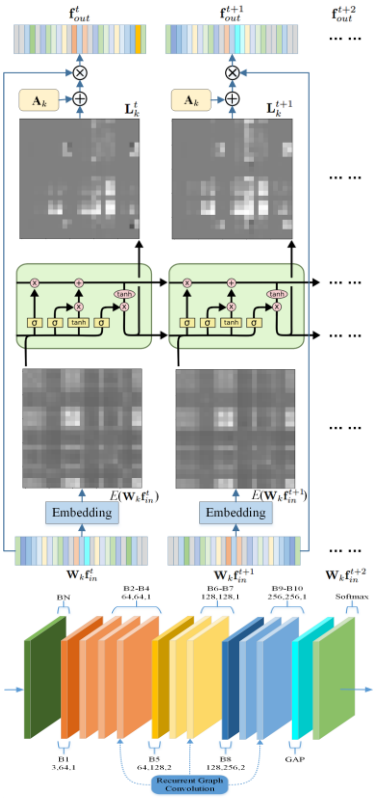
RESULTS

- We choose the "Dot Product" as the embedding function, since it needs the least computational cost and parameter size.
- Proposed R-GCN achieves state-of-the-art performances on the NTU RGB+D and Kinetics-Skeleton datasets, and outperforms 2s-AGCN in both cases.

DISCUSSION

- Employ LSTM for evolution of the graph topologies to learn interactions of different subsets of joints.

AMMO BAR



Networks	BP on Embedding	Natural Connectivity	Accuracy(%)	Parameter Size
R-GCN + Dot-Prod		√	93.8	3.2M
	√	√	93.5	
	√		**94.2**	
			93.8	
R-GCN + Gaussian		√	93.9	3.2M
	√	√	94.1	
R-GCN + EmbedGaussian		√	93.9	3.8M
	√	√	**94.2**	
R-GCN* + Dot-Prod		√	93.7	3.3M
	√		93.2	
	√	√	93.9	
	√		93.7	

Networks	Stream	NTU RGB+D		Kinetics-Skeleton	
		X-Sub(%)	X-View(%)	Top-1(%)	Top-5(%)
2s-AGCN [9]	Joint	-	93.2	35.1	57.1
	Bone	-	93.2	33.3	55.7
	Both	88.5	95.1	36.1	58.7
AGC-LSTM [21]	Joint	87.5	93.5	-	-
	Part	87.5	93.8	-	-
	Both	**89.2**	95.0	-	-
R-GCN	Joint	86.3	94.2	35.1	57.5
	Bone	87.6	94.1	34.6	57.1
	Both	**89.2**	**95.6**	**36.9**	**59.8**

Guangming Zhu, Lu Yang, Liang Zhang, Peiyi Shen and Juan Song

西安电子科技大学 XIDIAN UNIVERSITY

Take a picture to star our project in the GitHub.